

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2003, 08:45:46 ; Search time 25 Seconds  
(without alignments)  
744.915 Million cell updates/sec

Title: US-09-768-781-3

Perfect score: 2316

Sequence: 1 MDRVYIPEPNVDPPVSSLE.....RTRVENSEPPFETEARQSVV 449

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	925	39.9	444	1	XX HUMAN
2	191.5	8.3	159	1	XXRY HUMAN
3	106.5	4.6	497	1	DTPT LACHE
4	105.5	4.6	1584	1	BAIL HUMAN
5	104.5	4.5	741	1	YBIO ECOLI
6	102.5	4.4	405	1	YMP0 YEAST
7	102.5	4.4	521	1	YT25 CAEEL
8	100.5	4.3	264	1	ATP6 PODAN
9	100.5	4.3	459	1	N4M4M CERSI
10	99	4.3	233	1	ATP1 OCHNE
11	99	4.3	458	1	ACHO HUMAN
12	98.5	4.3	456	1	G64A DROME
13	98.5	4.3	684	1	TC10 YEAST
14	98.5	4.3	1033	1	YD56 SCHPO
15	97.5	4.2	325	1	MCSR RAT
16	97.5	4.2	551	1	VPH1 YEAST
17	96.5	4.2	805	1	YB81 SCHPO
18	96	4.1	406	1	HOF6 HAEIN
19	95.5	4.1	252	1	ATP6 NEUR
20	95.5	4.1	255	1	ATP6 YARLI
21	95.5	4.1	325	1	MCSR MOUSE
22	95	4.1	430	1	RFBX SALTY
23	95	4.1	692	1	NUSC MARPO
24	94.5	4.1	382	1	CKD6 RAT
25	94.5	4.1	438	1	MNT2 PSRAE
26	94.5	4.1	468	1	CB1A FUGRU
27	94.5	4.1	786	1	APE1 SULTO
28	94	4.1	607	1	NUSM MOUSE
29	94	4.1	715	1	LCCL LACIA
30	93.5	4.0	2376	1	YIM9 YEAST
31	92.5	4.0	371	1	NY6R YEAST
32	92.5	4.0	552	1	NUSM RHISA
33	92	4.0	248	1	ATPI MARPO

34 92 4.0 455 1 ACHO CHICK  
35 92 4.0 576 1 NUSM ANOQU  
36 91.5 4.0 378 1 CKR7 MOUSE  
37 91.5 4.0 429 1 SECY AQUAE  
38 91.5 4.0 464 1 NY4M PARLI  
39 91.5 4.0 628 1 YB1F SCHPO  
40 91 3.9 298 1 FIG1 YEAST  
41 91 3.9 394 1 G22A DROME  
42 91 3.9 1597 1 RLRI YEAST  
43 90.5 3.9 408 1 GS8B DROME  
44 90.5 3.9 567 1 DSBD CAMJE  
45 90 3.9 382 1 NU2M CHLRE

#### ALIGNMENTS

RESULT 1  
XX\_HUMAN STANDARD; PRT; 444 AA.  
AC PS1811;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Membrane transport protein XK (Kx antigen).  
GN XK.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94273191; PubMed=8004674;  
RA Ho M., Chelly J., Carter N., Danek A., Crocker P., Monaco A.P.;  
RT "Isolation of the gene for McLeod syndrome that encodes a novel  
RT membrane transport protein.";  
RL Cell 77:869-880(1994).  
RN [2]  
RP REVISIONS TO 204-205.  
RA Ho M.;  
Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: MAY BE INVOLVED IN SODIUM-DEPENDENT TRANSPORT OF NEUTRAL  
CC AMINO ACIDS OR OLIGOPETIDES.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- TISSUE SPECIFICITY: HIGH LEVELS IN SKELETAL MUSCLE, HEART, BRAIN,  
CC AND PANCREAS; LOW LEVELS IN PLACENTA, LUNG, LIVER, AND KIDNEY.  
CC -!- POLYMORPHISM: XK IS RESPONSIBLE FOR THE KX BLOOD GROUP SYSTEM.  
CC -!- DISEASE: DEFECTS IN XK ARE THE CAUSE OF MCLEOD SYNDROME, AN X-  
CC LINKED MULTISYSTEM DISORDER CHARACTERIZED BY ABNORMALITIES IN THE  
CC NEUROMUSCULAR AND HEMATOPOIETIC SYSTEMS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; Z32684; CAA83632.2; -;  
CC Genew; HGNC:12811; XK.  
CC MIM; 314850; -;  
CC Transmembrane; Transport; Amino-acid transport; Blood group antigen.  
CC DOMAIN 1 2  
CC TRANSMEM 3 23 POTENTIAL.  
CC TRANSMEM 24 37 EXTRACELLULAR (POTENTIAL).  
CC TRANSMEM 38 58 POTENTIAL.  
CC TRANSMEM 59 68 CYTOPLASMIC (POTENTIAL).  
CC TRANSMEM 69 89 POTENTIAL.  
CC TRANSMEM 90 140 EXTRACELLULAR (POTENTIAL).  
CC TRANSMEM 141 161 POTENTIAL.  
CC TRANSMEM 162 171 CYTOPLASMIC (POTENTIAL).  
CC TRANSMEM 172 192 POTENTIAL.

FT DOMAIN 193 208 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 209 229 POTENTIAL.  
FT DOMAIN 230 235 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 236 256 POTENTIAL.  
FT DOMAIN 257 277 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 278 298 POTENTIAL.  
FT DOMAIN 299 317 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 318 338 POTENTIAL.  
FT DOMAIN 339 349 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 350 370 POTENTIAL.  
FT DOMAIN 371 444 CYTOPLASMIC (POTENTIAL).  
SQ SEQUENCE 444 AA; 50902 MW; 6F90B0B45659A1DA CRC64;  
Query Match 39.9%; Score 925; DB 1; Length 444;  
Best Local Similarity 44.5%; Pred. No. 2.4e-60;  
Matches 179; Conservative 80; Mismatches 133; Indels 10; Gaps 4;  
QY 33 PPSILSTFLYCGEASALYVRIYKNSERYMTYTFSPFMFSIMVQLTLIFVHRDL 92  
DB 3 FPASVLASVPLFAETTAALSLSSTYRGDRMQALTLFSLPCALVQLTLFVHRDL 62  
QY 93 AKDKPLSLFHLILLGPVIRCLEAMIKYLTWKKEEQEPYVSLTRKK-MLIDGEEVLIE 151  
DB 63 SRDRPLVLLHLLQGLPFRFCFEVCIY---FQSGNNEEPYVSIITKQMPKQGLSEBIE 119  
QY 152 WEVCHSIRTLAMHRYKMSQIQAPLGSVQLTYQLVLSISARVPLGRVVLVFLSVLS 211  
DB 120 KEVGOAEGKLTHRSASFASRASVIOAPLGSAPQLTYQLVLSVQMDDVTYVGRSLMTISLS 179  
QY 212 VTGATLCNMLAIOIKYDDYKIRLGPVLVLCITITWRTLEITSRLILVLSATLKLKAVP 271  
DB 180 IVGALRCNLAIAIKIDEYEVKVKPLAYVIFLWRSFEIATRVVVLVFLTSVLTWVVV 239  
QY 272 FLVNLFIILPEPWIKFWRSGAQMNNIEKNFSRVGTLLVLSVTLVYAGINFSCWSALQ 331  
DB 240 IILNFFSFPLYPWLFWCSSGPPENIEKALSRYGTTIVLCFLTLTYTGINMFCWSAVQ 299  
QY 332 LRLADRLVDKGQWGHMGLHYSLVENVIMVLVFKFGVKVLLNCHSLIALQLIITAY 391  
DB 300 LKIDSPILSKSHNWQLLVYIMRIFENAILLWLYLFTKDIYVVCAPLLVLQLLIGY 359  
QY 392 LISIDFMLLFQYLHPLRSLSLFTNVVD---YLHCVC--CHQ 427  
DB 360 CTALFLMVFYQFHPCKKLSPSSVSSEGFQWLRCFCWACRQ 401

RESULT 2  
XKRY HUMAN STANDARD; PRT; 159 AA.  
AC O14609;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Testis-specific XK related protein Y.  
GN XKRY.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP MEDLINE=98022381; PubMed=9381176;  
RX Lahn B.T., Page D.C.;  
RT "Functional coherence of the human Y chromosome.";  
RL Science 278:675-680(1997).  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -1- TISSUE SPECIFICITY: TESTIS-SPECIFIC.  
CC -1- SIMILARITY: SOME, TO MEMBRANE TRANSPORT PROTEIN XK.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL: U77486; AAC45382.1; --  
CC InterPro: IPR000109; PIR2.  
CC Pfam: PF00854; PIR2; 1.  
CC TIGRFAMs: TIGR00923; 2A1701; 1.  
CC PROSITE: PS01022; PIR2.1; 1.  
CC PROSITE: PS01023; PIR2.2; 1.  
CC Peptide transport; Transport; Transmembrane.  
FT TRANSMEM 3 23 POTENTIAL.  
FT TRANSMEM 26 46

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL: AF000997; AAC51844.1; --  
CC DR GENE; HGNC:18571; XKRY.  
CC MIN; 400015; --  
CC Transmembrane. 1 21 POTENTIAL.  
CC TRANSMEM 45 65 POTENTIAL.  
CC TRANSMEM 72 92 POTENTIAL.  
CC SEQUENCE 159 AA; 18083 MW; B9E81DD842DEFOAB CRC64;  
Query Match 8.3%; Score 191.5; DB 1; Length 159;  
Best Local Similarity 31.2%; Pred. No. 1.9e-07;  
Matches 45; Conservative 39; Mismatches 51; Indels 9; Gaps 4;  
QY 206 VFSLSVTYGATLCNMLAIOIKYDDYKIRLGPVLVLCITITWRTLEITSRLILVLSATL 265  
DB 11 IFPLISCV-GAHCNLAIRTNDFPAIKQVILYLMHSLVLIISPVVTLAAPPASL 69  
QY 266 KLKAVPFLVNLFIILPEPWIKFWRSGAQMNNIEKNFSRVGTLLVLSVTLVYAGI--- 322  
DB 70 KQSLHFLIITYFVLLITPMLPEFSGKSGTHLPSTNK---IIPAWVWMDAYLNHASICCH 125  
QY 323 NFSCWSALQLRLADRLVDKQGNW 346  
DB 126 QFSLSAVKQLQSLNEELI-RDTRW 148  
RESULT 3  
DPTT LACHE STANDARD; PRT; 497 AA.  
ID DPTT LACHE STANDARD; PRT; 497 AA.  
AC O07380;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE DI-/tripeptide transporter.  
GN DPTT.  
OS Lactobacillus helveticus.  
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
OC Lactobacillus.  
OX NCBI\_TaxID=1587;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=NCDO 2712;  
RX MEDLINE=97316430; PubMed=9172341;  
RA Nakajima H., Hacting A., Kunji E.R.S., Poolman B., Konings W.N.;  
RT "Cloning and functional expression in Escherichia coli of the gene  
RT encoding the di- and tripeptide transport protein of Lactobacillus  
RT helveticus.";  
RL Appl. Environ. Microbiol. 63:2213-2217(1997).  
CC -1- FUNCTION: PROTON-DEPENDENT UPTAKE OF DI- OR TRI-PEPTIDES.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- SIMILARITY: BELONGS TO THE PIR2 FAMILY OF TRANSPORTERS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL: U77486; AAC45382.1; --  
CC InterPro: IPR000109; PIR2.  
CC Pfam: PF00854; PIR2; 1.  
CC TIGRFAMs: TIGR00923; 2A1701; 1.  
CC PROSITE: PS01022; PIR2.1; 1.  
CC PROSITE: PS01023; PIR2.2; 1.  
CC Peptide transport; Transport; Transmembrane.  
FT TRANSMEM 3 23 POTENTIAL.  
FT TRANSMEM 26 46

```
FT TRANSMEM 57 77 POTENTIAL.
FT TRANSMEM 84 104 POTENTIAL.
FT TRANSMEM 119 139 POTENTIAL.
FT TRANSMEM 155 175 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
FT TRANSMEM 227 247 POTENTIAL.
FT TRANSMEM 294 314 POTENTIAL.
FT TRANSMEM 321 341 POTENTIAL.
FT TRANSMEM 372 392 POTENTIAL.
FT TRANSMEM 452 472 POTENTIAL.
SQ SEQUENCE 497 AA; 55469 MW; 65C064FF9A0A26A5 CRC64;

Query Match 4.6%; Score 106.5; DB 1; Length 497;
Best Local Similarity 20.8%; Pred. No. 1;
Matches 85; Conservative 65; Mismatches 133; Indels 125; Gaps 18;

QY 91 DLAKDKPLSLFMHLLILGPVIR-----CLEAMI 118
Db 110 DPRDAGSIFVFGINLGSIIAPMLVPWAAQGVHIFGSQINFHAGFSLAAGVFFGLV 169
QY 119 KYLPLWKKEOEPEPVSLTRKMLIDGEEV-LIEWEVGHSIRTLAMHNRNAYKRWMSQIOAF 177
Db 170 QYVLGGKYLSTE---SLTPNDPKDGLNLVKKVVIITAIIVA-----ILAA 215
QY 178 LGSVPQTYQLYVSL---ISAEVPLGRVLMVFS--LVSVTYGATLCNM----- 221
Db 216 MAGVQLSDVNIVTLTTLATLPIYVPMFRSSKVTKIELGIHLPLVSLKNRLFPPKKG 275
QY 222 -----LAIQIKYDYDKIRLGPVLCITWRTLEITRLLILVLF-SATLKLKAVP 271
Db 276 YKRLKQIQLLELAIKROSFIILAIIMASILIPNKVIAKHLKLVLVYFVIGLNLIP 335
QY 272 F--LVNLFLIILPEPWIK--PWRSGAQPNNIEKN-----FSR-----VGTLLVVL 312
Db 336 FSTFVLFS--LFDYIKHMKFKGEQAKTKYKSRHHGIEIPLFLQLIINIFTLIL 392
QY 313 ISVTIL-YAGINFSC-----WSALQLRLADRLVDKQNGHMGHLYSVRLVENIMV 364
Db 393 EGETLPDENGVEVNIABHPVQGYTELINLANKSIDLWADW-----IQSV--- 438
QY 365 LVPKFFGVKVLNLYCHSLALQLLIAYLISIDFMLLPQYVLPRLSLF 412
Db 439 -----AKYLLNIMYADVIVIIIFVLYKMAALWMAWSYI-PLSTVF 478

RESULT 4
BAIL_HUMAN STANDARD; PRT; 1584 AA.
AC O14514;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Brain-specific angiogenesis inhibitor 1 precursor.
GN BAIL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=98054121; PubMed=9393972;
RA Nishimori H., Shiratsuchi T., Urano T., Kimura Y., Kiyono K.,
RA Tatsumi K., Yoshida S., Ono M., Kuwano M., Nakamura Y., Tokino T.;
RT "A novel brain-specific p53-target gene, BAIL, containing
RT thrombospondin type 1 repeats inhibits experimental angiogenesis.";
RL Oncogene 15:2145-2150(1997).
RN [2]
RP INTERACTION WITH BAP1.
RX MEDLINE=98321173; PubMed=9647739;
RA Shiratsuchi T., Futamura M., Oda K., Nishimori H., Nakamura Y.,
RA Tokino T.;
RT "Cloning and characterization of BAI-associated protein 1: a PDZ
```

```
RT domain-containing protein that interacts with BAIL.";
RL Biochem. Biophys. Res. Commun. 247:597-604(1998).
CC -!- FUNCTION: LIKELY TO BE A POTENT INHIBITOR OF ANGIOGENESIS IN
CC BRAIN AND MAY PLAY A SIGNIFICANT ROLE AS A MEDIATOR OF THE P53
CC SIGNAL IN SUPPRESSION OF GLOBLASTOMA. MAY FUNCTION IN CELL
CC ADHESION AND SIGNAL TRANSDUCTION IN THE BRAIN.
CC -!- SUBUNIT: INTERACTS WITH BAP1.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. LIKELY TO BE
CC CONCENTRATED AT CELL-CELL ADHESION SITES.
CC -!- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN BRAIN. REDUCED OR NO
CC EXPRESSION IS OBSERVED IN SOME GLOBLASTOMA CELL LINES AND CANCER
CC TISSUES.
CC -!- INDUCTION: BY P53.
CC -!- DOMAIN: THE TSPI REPEATS INHIBIT IN VIVO ANGIOGENESIS IN RAT
CC CORNEA INDUCED BY BFGF.
CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -!- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 GPS DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB005297; BAA23647.1; -.
DR Genbank; HGNC:943; BAIL.
DR MIM; 602682; -.
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR000203; PKD_cys_rich.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR001879; hormn_receptor.
DR Pfam; PF00002; 7tm_2; 1.
DR Pfam; PF00090; tsp_1; 5.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00008; Hormr; 1.
DR SMART; SM00209; TSPI; 5.
DR PROSITE; PS00221; GPS; 1.
DR PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
DR PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
DR PROSITE; PS00227; G_PROTEIN_RECEP_F2_3; 1.
DR PROSITE; PS00263; G_PROTEIN_RECEP_F2_4; 1.
DR PROSITE; PS00092; TSPI; 5.
DR G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
DR Repeat; Cell adhesion.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1584 BRAIN-SPECIFIC ANGIOGENESIS INHIBITOR 1.
FT DOMAIN 31 948 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 949 969 1 (POTENTIAL).
FT DOMAIN 970 980 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 981 1001 2 (POTENTIAL).
FT DOMAIN 1002 1008 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1009 1029 3 (POTENTIAL).
FT DOMAIN 1030 1052 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1053 1073 4 (POTENTIAL).
FT DOMAIN 1074 1093 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1094 1114 5 (POTENTIAL).
FT DOMAIN 1115 1136 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1137 1157 6 (POTENTIAL).
FT DOMAIN 1158 1166 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1167 1187 7 (POTENTIAL).
FT DOMAIN 1188 1584 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 261 316 TSP TYPE-1 1.
FT DOMAIN 354 408 TSP TYPE-1 2.
FT TRANSMEM 409 463 TSP TYPE-1 3.
FT DOMAIN 467 521 TSP TYPE-1 4.
FT TRANSMEM 522 576 TSP TYPE-1 5.
FT DOMAIN 581 938 GPS.
FT TRANSMEM 1411 1422 POLY-PRO.
```

```

DNA Res. 3:137-155(1996).
RL  -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC  -1- SIMILARITY: BELONGS TO THE UPF0003 FAMILY.
CC  -1- CAUTION: Ref.2 sequence differs from that shown due to a
CC      frameshift in position 77.
CC -----
CC  This swiss-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC -----
CC  ENBL; AE000183; AAC73895.1; ALT_INIT.
CC  DR  ENBL; D90717; BAA35474.1; -.
CC  DR  ENBL; D90718; BAA35479.1; ALT_FRAME.
CC  DR  ENBL; D90719; BAA35480.1; ALT_FRAME.
CC  DR  EcoGene; EGI3320; ybio.
CC  DR  InterPro; IPR001880; MSion_channel.
CC  DR  Pfam; PF00924; MS_channel_1.
CC  DR  PROSITE; PS01246; UPF0003; FALSE_NEG.
CC  KW  Hypothetical protein; Transmembrane; Complete proteome.
FT  TRANSMEM 9 29
FT  TRANSMEM 143 163
FT  TRANSMEM 185 205
FT  TRANSMEM 225 245
FT  TRANSMEM 268 288
FT  TRANSMEM 294 314
FT  TRANSMEM 343 363
FT  TRANSMEM 374 394
FT  TRANSMEM 432 452
FT  TRANSMEM 466 486
FT  TRANSMEM 509 529
FT  TRANSMEM 533 553
FT  TRANSMEM 562 582
FT  TRANSMEM 608 628
FT  TRANSMEM 662 682
FT  DOMAIN 25 30
FT  DOMAIN 195 202
FT  POLY-LEU.
SQ  SEQUENCE 741 AA; 81903 MW; 69B0A27678976DB5 CRC64;

Query Match
Best Local Similarity 19.5%; Score 104.5; DB 1; Length 741;
Matches 96; Conservative 91; Mismatches 173; Indels 133; Gaps 25;

Qy 1 MDRVETPEEPNDVPVSSLEEDVIRGANPRTFPFSLFSTFL-----YCEAASAL 52
Db 1::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
66 IDQRTVAATPPAPPV-----PRIVPTLVEEQTLQKVTEVSRHYGALSAR 113

Qy 53 YMVRIYRK-----NSETRYMTV-----FSFMFSSIMVQLTLIFVHRLA- 93
Db 114 F-GOLYRNITGSPHKPNPOTFSNALTFHSMALVIVGFYV----LRLCALPYRKMGQ 168

Qy 94 ----KQKPLSLFMHL--ILLGPVIRCLAEAMIKYTLWKKEQEEPPYSLTRKMLDGE 147
Db 1::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
169 WARQKNRERNWLQLPAMIIGAFI--IDLLALLALTFLVQG----- 206

Qy 148 VLIEWEVGHISIRTLAMERNAYKRMSQIQAFLGSPVQITY-----QLVSLISAEP--LG 200
Db 1::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
207 -LVSDNLNAGSRITIAFOQSLSFLNAFALIEFPKAVLRILFCPNVAELRPFPTIQDESARYWS 265

Qy 201 RVLVMFSLSVYVYGATLCNMLAIQIKYDDYDKYRLGPLE---VLCITIW-----RTL 249
Db 1::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
266 RRLSWLSSLIG--YGL-----IVAVPIISNQVNVQIGALANVIMLCVTWVLYLFRNKK 319

Qy 250 EITSRL-----IIVLFSATLK-----LKAVPFFVLVNLFLIILFPFWIKFWRSGAQP 296
Db 1::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
320 EITQHLNFAEHSLAFSLFIRAFALVWHMLASAYFIVL-FFFSLFDP-----G 367

Qy 297 NNTKFNFRVGTLLVLLISVTILVAGINFCWSALQLRLADR-----DLVDKQNWGHMG 350
Db 1::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
368 NSLKFMFGATVRSIIAIGIAAFVSGM-FSWLAKTITLSPHQRYNPPELOKRLNGWLSAA 426

```

```

FT IRANSHEM 682 682 POLY-THR.
FT DOMAIN 25 POLY-LEU.
FT DOMAIN 195 202 POLY-LEU.
SQ SEQUENCE 741 AA; 81903 MW; 69B0A27678976DB5 CRC64;

Query Match 4.5%; Score 104.5; DB 1; Length 741;
Best Local Similarity 19.3%; Pred. No. 2.3;
Matches 96; Conservative 91; Mismatches 173; Indels 133; Gaps 25;

Qy 1 MDRVYEPEEPNVDVSSLEEDVIRGANPFTPEFSLFSTFL-----YCGEASAL 52
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 IDQURTVAAETPAEPV-----PKIVPTLVSEQTVLOKVTEVSRHYGEALSAR 113
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 53 YMVRIYRK-----NSETYRMTYT-----ESFFMFSSIMVQLTLIFVHRDLA- 93
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 114 F-GOLYXRNITGSPHKPNPQTFSNALTFESMLAVLVFGFYW---LIRLCALPLYRXMGQ 168
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 94 ----KQPLSLFPHL---ILGPIVRCLEAMKIVTLWKKEEQEBEYVSLTRKKMLDGE 147
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 169 WAROKNRSRNNWLQLPAMIICAFI-IDLLALLATLVFGQ-----206
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 148 VLIWEVGHGSIRTLAMERNAYKMSQIQALFSGVPOLTY-----QLVSLISAEPV--LG 200
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 207 -VLVDNLNAGSRTIAFOOSLFLNAPALIEPFKAVLRILFCFNVAEALRPPTIQDESARYWS 265
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 201 RVLMVPSLVSVTYGATLCNMLAIQIKYDDYDKIRLGPLE---VLCITIW-----RTL 249
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 266 RRLSWLSSLTG--YGL-----IVAVPIISQNVQVIGALANVIIMLCWTWALYLI FRNKK 319
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 250 EITSRL-----ILVLFSA TLK-----LKAVPFLVLNPLLIILFPFWIKFWRSGAQP 296
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 320 EITQHLINFAEHSLAPSLFIRAFALVWHMLASAYFIVL-FFFSLFDP-----G 367
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 297 NNTKXFSRVGTLVLLISVTILVAGINFCWSALQLRLADR-----DLVDKQGNWGHMG 350
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 368 NSLKFMFGATVRSIIAIGIAAFVSGM-FSRWLAKTITLSPHTRQNYPELOKRLNGWLSAA 426
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```



```
Db      162 -----IAUTHIQNLGTSLNLL---IQVSTQTPSSWSNAFLWLACMM 201
Oy      176 AFLGSVPQLTYQLVWSLISAEVL-GRVLWMFSLVSVTVGATLCNNLAIQIKYDDYKIR 234
```

```

Db 202 AFVVKMPYGLHLWLPKARVEAPIAGSNVLAAILLKGYGMLRTMI----- 249
Qy 235 LGPL-----EVLCTITWRTLEITSRLILVLPATLKLKAVPFLVNLFLILFE-PWIK 287
Db 250 LNPITSYMAYPFLMLSLWGMIMTSSICLRQDLSLAYSSVSHMALVIVAVLIOTPW-S 308
Qy 288 FWRSGAQM-----PNNIEKNFSR-----VGTLVVLISVTILYAGINFSC 326
Db 309 YMGATALMIAHGLTSSVLFCLANSYERTHRTWILARGLOTLPAMWMLASLT--- 365
Qy 327 WSALQRLADRLVDKQNGHGMGLHYSVRLENVIMVL-VKFFGKVLVNLNCHSLIAL 385
Db 366 -----NLALPPTINLVGELFVMSFSNSNITII-----LMGT 398
Qy 386 QLIIVLISIDFMLFFOYHPLRLSTHNVDVLCVCHQHQRTRVENSEPPFETE 443
Db 399 NIITIALYSL-YMLITTO-----RGKYTHH-----INNPKPSFTRE 433

RESULT 10
ATPI_OCHNE
ID _ATPI_OCHNE STANDARD; PRT; 233 AA.
AC Q40607;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP synthase A chain precursor (EC 3.6.3.14) (Subunit IV).
GN ATPI.
OS Ochrosphaera neapolitana.
OC Chloroplast.
OC Eukaryota; Haptophyceae; Isochrysidales; Ochrosphaera.
OX NCBI_TaxID=35137;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CCMP 593;
RA Huss V.A.R., Tietze A.C., Julius C.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
CC DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +
CC H(+) (out).
CC -1- SUBUNIT: P-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
CC thylakoid membrane.
CC -1- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X99078; CAA67535.1; -
CC InterPro; IPR000568; ATPsynT_Asub.
CC Pfam; PF00119; ATP-synt_A; 1.
CC TIGRFAMs; TIGR01131; ATP_synt_6_or_A; 1.
CC PROSITE; PS00449; ATPASE_A; 1.
CC Hydrogen ion transport; CF(0); Chloroplast; Transmembrane; Signal.
CC SIGNAL 1 ? BY SIMILARITY.
CC CHAIN ? 233 ATP SYNTHASE A CHAIN.
CC FT TRANSMEM 27 47 POTENTIAL.
CC FT TRANSMEM 84 104 POTENTIAL.
CC FT TRANSMEM 192 212 POTENTIAL.
CC SEQUENCE 233 AA; 25721 MW; 6EEA74DE0887D6F1 CRC64;

Query Match 4.38; Score 99; DB 1; Length 233;
Best Local Similarity 23.74; Pred. No. 1.6;

```

```

Matches 64; Conservative 22; Mismatches 68; Indels 116; Gaps 13;
Qy 205 MVSFSLVSVYTGATLCNMLAIOIKYDDYKIRLGPLE-----VLCITWRTLEITSRLILVL 260
Db 1 MFFSLAAVEVTHL-----YWEIGGLEVHGQVLLIT-WLVLAAILTLAIL-- 44
Qy 261 FSATLKLKAVPFLVNLFLILFE-----PWIKF-----WRSQ 292
Db 45 --GTLKLEQVFKQNFLESFEYVSGIAKQDIGYHYRPMVPFVGTFLFIFVANW-LG 101
Qy 293 AQMP-----NNIEKNFSRVGTILVVLISVTILYAGINFSCWSALQLRLADR 337
Db 102 ALIPKWLITLPEGLAAPTNDINTTVA-----LSLTSISYFYAGFK----- 143
Qy 338 DLVDKQNGHGMGLHYSVRLVE-----NVI-----MVLVRFKFFG-----VKVL 375
Db 144 -----EKGIGFFARYISPTPIFLPINILEDTFKPLSLFRFLGNILADEIVVSVL 193
Qy 376 LNYCHSLIALQLIIAYLISIDFMLFFOYL 405
Db 194 CLLVPLLIPLPVMVLGIFPASSVQALVFSTL 223

RESULT 11
ACHO_HUMAN
ID _ACHO_HUMAN STANDARD; PRT; 458 AA.
AC Q05901; Q15827;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neuronal acetylcholine receptor protein, beta-3 chain precursor.
GN CHRN3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Substantia nigra;
RX MEDLINE=97062879; PubMed=8906617;
RA Elliott K.J., Ellis S.B., Berckhan K.J., Urrutia A.,
RA Chavez-Noriega L.E., Johnson E.C., Velicelebi G., Harpold M.M.;
RT "Comparative structure of human neuronal alpha 2-alpha 7 and beta
RT 2-beta 4 nicotinic acetylcholine receptor subunits and functional
RT expression of the alpha 2, alpha 3, alpha 4, alpha 7, beta 2, and
RT beta 4 subunits.";
RL J. Mol. Neurosci. 7:217-228(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pons;
RX MEDLINE=97162233; PubMed=9009220;
RA Groot Kormelink P.J., Luyten W.H.M.L.;
RT "Cloning and sequence of full-length cDNAs encoding the human neuronal
RT nicotinic acetylcholine receptor (nAChR) subunits beta3 and beta4 and
RT expression of seven nAChR subunits in the human neuroblastoma cell
RT line SH-SY5 and/or IMR-32.";
RL FEBS Lett. 400:309-314(1997).
RN [3]
RP SEQUENCE FROM N.A.
RA Keddache M., Durner M., Greenberg D.A.;
RT "Genomic structure and mutation analysis of the CHRN3 gene.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 36-458 FROM N.A.
RC TISSUE=Brain stem;
RX MEDLINE=9339761; PubMed=7690916;
RA Willoughby J.J., Ninkina N.N., Beech M.M., Latchman D.S., Wood J.N.;
RT "Molecular cloning of a human neuronal nicotinic acetylcholine
RT receptor beta 3-like subunit.";
RL Neurosci. Lett. 155:136-139(1993).
CC -1- FUNCTION: AFTER BINDING ACETYLCHOLINE, THE ACHR RESPONDS BY AN
CC EXTENSIVE CHANGE IN CONFORMATION THAT AFFECTS ALL SUBUNITS AND
CC LEADS TO OPENING OF AN ION-CONDUCTING CHANNEL ACROSS THE PLASMA

```

```

CC CC MEMBRANE.
CC CC -1- SUBUNIT: NEURONAL ACHR SEEMS TO BE COMPOSED OF TWO DIFFERENT
CC CC TYPE OF SUBUNITS: ALPHA AND BETA.
CC CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNEL FAMILY.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch)
CC CC
CC CC EMBL; U62438; AAB40116.1; -.
CC CC EMBL; Y08417; CRA69694.1; -.
CC CC EMBL; AF140765; AAD33063.1; -.
CC CC EMBL; AF140765; AAD33063.1; JOINED.
CC CC EMBL; AF140761; AAD33063.1; JOINED.
CC CC EMBL; AF140762; AAD33063.1; JOINED.
CC CC EMBL; AF140763; AAD33063.1; JOINED.
CC CC EMBL; AF140764; AAD33063.1; JOINED.
CC CC EMBL; X67513; CAA47851.1; -.
CC CC PIR; S25587; S25587.
CC CC Genew; HGNC:1963; CHRN3.
CC CC
CC CC MIM; 118508; -.
CC CC InterPro; IPR000188; GABAA receptor.
CC CC InterPro; IPR001175; Neur_channel.
CC CC Pfam; PF02931; Neur_chan_LBD; 1.
CC CC Pfam; PF02932; Neur_chan_memb; 1.
CC CC PRINTS; PR00252; NRIONCHANNEL.
CC CC TIGRFAMs; TIGR00860; LIC; 1.
CC CC PROSITE; PS00236; NEUROTR_ION_CHANNEL; 1.
CC CC Postsynaptic membrane; Ionic channel; Glycoprotein; Signal;
CC CC Transmembrane; Multigene family.
CC CC SIGNAL 1 24
CC CC CHAIN 25 458
CC CC
CC CC DOMAIN 25 232
CC CC TRANSMEM 233 257
CC CC TRANSMEM 265 282
CC CC TRANSMEM 299 320
CC CC DOMAIN 321 428
CC CC TRANSMEM 429 447
CC CC DISULFD 153 167
CC CC CARBOHYD 51 51
CC CC CARBOHYD 138 138
CC CC CARBOHYD 166 166
CC CC CONFLICT 36 38
CC CC SEQUENCE 458 AA; 52728 MW; D6E919E53CBD21F8 CRC64;
CC CC
CC CC Query Match 4.3%; Score 99; DB 1; Length 458;
CC CC Best Local Similarity 25.7%; Pred. No. 3.3;
CC CC Matches 45; Conservative 35; Mismatches 63; Indels 32; Gaps 8;
CC CC
CC CC Qy 73 PFMESSIMVQLTLLI---FV---HRDLAKDKPLSPMLHLLG--PVIRCLEAMIKYLLT 123
CC CC Db 299 YLLFTMIPTLLSIITVTVINVHRRSSSTTHPMAPVKKLFQKLPKLCMKDHDVDRYS 358
CC CC
CC CC Qy 124 WKKEQEPEPVY-----SLTRKMKLIDGEEVLIEW--EVGHSIRTLAMHNRNAYKMSQIOA 176
CC CC Db 359 PEKEE-SQPVVKGVLEKKQKQSDGKVLAFLEKAADSIRYSRHKKEHFTISQVVQ 417
CC CC
CC CC Qy 177 FLGSPVQTLTYQLYVLSIAEVPVGRVLMVFLSVVTVYATLGNMLAIGIKYDDY 231
CC CC Db 418 DWKFVAQV-----LDRIFLFLFIVSVT-GSVLIPTALKMWLHYS 457
CC CC
CC CC RESULT 12
CC CC G64A DROME STANDARD; PRT; 456 AA.
CC CC AC P83293; Q9VZJ7;
CC CC DT 15-JUN-2002 (Rel. 41, Created)

```

```

DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative gustatory receptor 64a.
GN GR64A OR CG14986.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Frannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Glisler A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RN IDENTIFICATION.
RX MEDLINE=21407712; PubMed=11516643;
RA Dunipace L., Meister S., McNealy C., Amrein H.;
RT "Spatially restricted expression of candidate taste receptors in the
RT Drosophila gustatory system.";
RL Curr. Biol. 11:822-835(2001).
RN [3]
RN CONCEPTUAL TRANSLATION.
RA Robertson H.;
RL Unpublished observations (NOV-2001).
CC -1- FUNCTION: Probable role in the gustatory response.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO FAMILY DR-TR OF G-PROTEIN COUPLED
CC RECEPTORS. SUBFAMILY II.
CC -1- CAUTION: Ref.1 sequence differs from that shown due to erroneous
CC gene model prediction.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```



modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; A5003480; AAF47824.1; ALT\_SEQ.  
 DR FlyBase; FBgn0045479; Gr64a.  
 KW Hypothetical protein; Receptor; G-protein coupled receptor;  
 Transmembrane; Glycoprotein; Multigene family.  
 FT DOMAIN 1 61  
 FT TRANSMEM 62 82  
 FT DOMAIN 83 95  
 FT TRANSMEM 96 116  
 FT DOMAIN 117 128  
 FT TRANSMEM 129 149  
 FT DOMAIN 150 185  
 FT TRANSMEM 186 206  
 FT DOMAIN 207 238  
 FT TRANSMEM 239 259  
 FT DOMAIN 260 321  
 FT TRANSMEM 322 342  
 FT DOMAIN 343 351  
 FT TRANSMEM 352 372  
 FT DOMAIN 373 456  
 FT CARBOHYD 52 52  
 FT CARBOHYD 217 217  
 SQ SEQUENCE 456 AA; 52822 MW; 5466974C73495D6C CRC64;

Query Match 4.3%; Score 98.5; DB 1; Length 456;  
 Best Local Similarity 20.7%; Pred. No. 3.6;  
 Matches 75; Conservative 60; Mismatches 117; Indels 111; Gaps 21;

QY 8 PEPNPDPV--SSLEEDVIRGAN-PRT-----PPSILFSTFLYCGEASALVMV 55  
 DB 27 PETPPPKFVEDSNLEFNVLASEKLPNTNLDLPHRAVPP-----FMFLAQCVAIMPLV 79  
 QY 56 RIYKNSSETMYT-TFSFFMFSSIMVQLTLIFVHRDLAKDKPLSLFMHLILGLPVIR-- 112  
 DB 80 GIRESNPRVRFAKSPIMFVTLFIATISLF-----LSMFTHLLKIGITAKNF 129  
 QY 113 -----CLEMIKYLTL-----WKKEB-----QEEPYVSLTRKMLIDGEE 147  
 DB 130 VGLVFFGCVLSAYVVFVIRLAKKPAVVRWIRTEIPETKPPYIPKRNLSRVQL---AA 186  
 QY 148 VLTEWEVGHISIRTLAMHR---NAYKMSQIOAFGLGVP-----OLTYQLYVLSIAEV 197  
 DB 187 LAI---IGLSGEHALYQVSAILSYTRRIQCANITVPSFNNWQNTYDVFOLL----- 239  
 QY 198 PLGRVLMVFLSVSVTYGATLCNNL-----AIQIKYDDYKIRLGPL--EVLIC 242  
 DB 240 PYSPIIAVLLIN--GA--CTFVWYMDLFIMWISKLSYRFEQITTRIKLEHEEVC 294  
 QY 243 ITTW-----RTLE-ITSLILLVLPSATLKUKAVPFLVNFILILFEP--WIKF 288  
 DB 295 ESFVIQIREHYVKMCELLEFVDSAMSSILLSCVNNLYFVCYQLNFKLRWPNVIYF 354  
 QY 289 WRS 291  
 DB 355 WYS 357

RESULT 13  
 TC10\_YEAST STANDARD; PRT; 684 AA.  
 ID TC10\_YEAST AC P50273;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE TC10 protein.  
 GN TC10 OR YDR350C OR D9476.9.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;

RN SEQUENCE FROM N.A.  
 RP STRAIN=MH125;  
 RC Zhang Y., Robinson K.M., Lemire B.D.;  
 RA Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
 RA Favellio A., Fullon L., Gattung S., Greco T., Kirsten J.,  
 RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,  
 RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,  
 RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,  
 RA Rifken L., Riles L., Taich A., Trevasakis E., Vignati D.,  
 RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.;  
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U32306; AAA74031.1; -;  
 DR EMBL; U28372; AAB64786.1; -;  
 DR SGD; S0002758; TCM10.  
 DR CONFLICT 592 684  
 FT GARSWYNKILFGGEIRHMALIOIKDOGWPPKPNFDETL  
 FT TELVENNNIKETPTSTLTFTDEDMFEEDGKPRNDODVNKCT  
 FT NIIRETLKSLN -> EHAPGTRIVSSGALKSGIWL (IN  
 FT REF. 2).  
 SQ SEQUENCE 684 AA; 79755 MW; A88992848F8F49A4 CRC64;

Query Match 4.3%; Score 98.5; DB 1; Length 684;  
 Best Local Similarity 18.7%; Pred. No. 5.7;  
 Matches 90; Conservative 60; Mismatches 163; Indels 169; Gaps 19;

QY 31 FTTPFPSTFLYCGEASALVMVRIYKNSSETMYTTFSTFFM----- 75  
 DB 173 FLKSDVLLFTSNCTFTNRLI-----KGTMERQATQIFLHDETNIKFIMEKVUK 226  
 QY 76 -----FSSIMVQLTL-----IFVHRDLAKDKPL-----SLF 101  
 DB 227 LHTFDSLIALVNLGVAKNPFKIVFIQALLQLEQHCYSGKDGAKQKNLRYVKFNNTLL 286  
 QY 102 MHLILLGPVIRCLAMIKYLTWKKEQEEPYVSLTRKMLIDGEEVLEWEGHSIRTL 161  
 DB 287 YLLKSGNV---ELFIKTF---QEEKFIYSSGLLNHIDGNEHILNFPPIHHYLNLL 336  
 QY 162 AMHENVAYKMSQIOAFGLGVPOLTYQLYVLSIAEVLGRVVMVFLSVSVTYGATLCNM 221  
 DB 337 RI-SNROEELFNVISCLQSSPLMKYKLPKPELMGE-----LIASFOAFRDPKLVCKY 387  
 QY 222 LAIQIKYDDYKIRLGPLEVLCITWRTL-EITSRLILLVLSATLKLKAVPELVNLFLLI 280  
 DB 388 L-----LSSYSKASANILNALGIWGLYHKSSTLTAPTARELNK----- 430  
 QY 281 LFEPWIKFWRSGAQMNNIEKNFSGVGLVLLISVTILYAGINFCWSALQRLADRLV 340  
 DB 431 -----NNILPNTWRIGSPVTPILTLYRSLLSS--SSVSL----- 464  
 QY 341 DKQNGWGHMGLHYSVRLVENVMVLFKFFG-----VKLLNVCYHSL 382  
 DB 465 ESGQ-----FKNCLDLYYKYSFLSEAHKYRYWRNDTGILNVLNLYIR-F 510  
 QY 383 IALQLIITAYLISIDFMLLFFOYLHPL-----RSLEFHNVDVYLHVCCHQHPTRVEN 435  
 DB 511 QAREPLAYNVLLDF-----YSQFFAKKVVLTTLTCLPFSIVAY-----KNHLTQASL 558  
 QY 436 SE 437  
 DB 559 SE 560

```
RESULT 14
YD56 SCHPO STANDARD; PRT; 1033 AA.
AC Q10309;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Potential phospholipid-transporting ATPase C6C3.06C (EC 3.6.3.1).
GN SPAC6C3.06C
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shipakovski G.V., Uesery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002)
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
(CC (E1-E2 ATPASES). SUBFAMILY IV.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; Z69731; CAA93618.1; -
CC InterPro; IPR001757; ATPase_E1-E2.
CC InterPro; IPR001454; Hlgnaase/hydrolase.
CC Pfam; PF00702; Hydrolase; 1.
CC PROSITE; PS00154; ATPase_E1_E2; 1.
KW Hypothetical protein; Hydrolase; Transmembrane; Phosphorylation;
KW Magnesium; ATP-binding.
FT TRANSMEM 133 153 POTENTIAL.
FT TRANSMEM 274 294 POTENTIAL.
FT TRANSMEM 317 337 POTENTIAL.
FT TRANSMEM 344 364 POTENTIAL.
FT TRANSMEM 768 788 POTENTIAL.
FT TRANSMEM 843 863 POTENTIAL.
FT TRANSMEM 913 933 POTENTIAL.
FT TRANSMEM 939 959 POTENTIAL.
-----
FT TRANSMEM 965 985 POTENTIAL.
FT TRANSMEM 992 1012 POTENTIAL.
FT MOD_RES 408 408 PHOSPHORYLATION (BY SIMILARITY).
FT METAL 770 770 MAGNESIUM (BY SIMILARITY).
FT METAL 774 774 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 1033 AA; 116587 MW; D57C467427D0C6D6 CRC64;
Query Match 4.3%; Score 98.5; DB 1; Length 1033;
Best Local Similarity 19.8%; Pred. No. 9;
Matches 62; Conservative 55; Mismatches 91; Indels 105; Gaps 15;
QY 46 GEAAS--ALYVRIYRNSE-----TYRMTYTFSPFMFSIMVQLTLIFVHRLAKD 95
DB 792 GQAASLAADYSVKESHSVRLLLWHGRISYKQTSKLAHMFVHGRLLISVCQVYSVISAF 851
QY 96 KPLSLFMHLIILGPVIRCLEAMIKYLTWKKEEQEPEPVSLTRKMLDGEVLEWEVG 155
DB 852 EPIALFOGLLLVG-----YSTM-----YTMLEFVSIVYDRDV----- 883
QY 156 HSIRTLAMHRNAYKMSQIOAFLGSPQLTYOLYVSLISAEPVPLGRVV-LMWFSLV---- 210
DB 884 -SEKLVFLFPELYKEMREKCF-----SYKNFISCVLSIVYQGLIQLFTFYLGFEF 935
QY 211 ----SVTYGATLCN--MLAIQIKYDDYKIRLGPLEVLCITIRWTLTITSRLIILVLF 262
DB 936 EGKMLAVCFSCLIFFNELIMVALQ-----INTWEQTIWVSELTLMY- 977
QY 263 ATCLKAVFVLVNLPLIILFSPWIKF-WRSGAQMNNEKNFSRVGTLLVLSIVTILYAG 321
DB 978 ----ILSVPEFLT-NYFELKFLGLKFWY-----VSALIFLISLLPVM-- 1014
QY 322 INFSCWSALQRL 334
DB 1015 ----CGKALKRKL 1023
RESULT 15
MC5R RAT STANDARD; PRT; 325 AA.
AC P35345;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Melanocortin-5 receptor (MC5-R).
GN MC5R.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Striatum;
RX MEDLINE=94234987; PubMed=8179577;
RA Griffon N., Mignon V., Facchinetti P., Diaz J., Schwartz J.C.,
RA Sokoloff P.;
RT "Molecular cloning and characterization of the rat fifth melanocortin
RT receptor."
RL Biochem. Biophys. Res. Commun. 200:1007-1014(1994).
CC -1- FUNCTION: RECEPTOR FOR MSH (ALPHA, BETA AND GAMMA) AND ACTH. THE
CC ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
CC ADENYLATE CYCLASE. THIS RECEPTOR IS A POSSIBLE MEDIATOR OF THE
CC IMMUNOMODULATION PROPERTIES OF MELANOCORTINS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: VERY LOW EXPRESSION LEVELS IS DETECTED IN
CC BRAIN, WHILE HIGH LEVELS ARE FOUND IN ADRENALS, STOMACH, LUNG AND
CC SPLEEN.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L27081; AAA41577.1; -.  
 PIR; JC2193; JC2193.  
 InterPro; IPR000276; GPCR\_Rhodospn.  
 Pfam; PF00001; 7tm\_1; 1.  
 PROSITE; PS00237; G\_PROTEIN\_RECF\_F1\_1; 1.  
 PROSITE; PS0362; G\_PROTEIN\_RECF\_F1\_2; 1.  
 G-protein coupled receptor; Transmembrane; Glycoprotein;  
 Phosphorylation; Lipoprotein; Palmitate.  
 DOMAIN 1 37 EXTRACELLULAR (POTENTIAL).  
 TRANSMEM 38 61 1 (POTENTIAL).  
 DOMAIN 62 73 CYTOPLASMIC (POTENTIAL).  
 TRANSMEM 74 97 2 (POTENTIAL).  
 DOMAIN 98 114 EXTRACELLULAR (POTENTIAL).  
 TRANSMEM 115 138 3 (POTENTIAL).  
 DOMAIN 139 155 CYTOPLASMIC (POTENTIAL).  
 TRANSMEM 156 179 4 (POTENTIAL).  
 DOMAIN 180 186 EXTRACELLULAR (POTENTIAL).  
 TRANSMEM 187 211 5 (POTENTIAL).  
 DOMAIN 212 239 CYTOPLASMIC (POTENTIAL).  
 TRANSMEM 240 265 6 (POTENTIAL).  
 DOMAIN 266 273 EXTRACELLULAR (POTENTIAL).  
 TRANSMEM 274 297 7 (POTENTIAL).  
 DOMAIN 298 325 CYTOPLASMIC (POTENTIAL).  
 CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).  
 LIPID PALMITATE (POTENTIAL).  
 LIPID PALMITATE (POTENTIAL).  
 SEQUENCE 325 AA; 37050 MW; 4E19FF1ABE8A6BDC CRC64;

Search completed: April 1, 2003, 08:47:43  
Job time : 30 secs